GlassBoARd: A Gaze-Enabled AR Interface for Collaborative Work

Kenan Bektaş
kenan.bektas@unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Adrian Pandjaitan
adrian.pandjaitan@student.unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Jannis Strecker
jannisrene.strecker@unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Simon Mayer
simon.mayer@unisg.ch
University of St. Gallen
St.Gallen, Switzerland

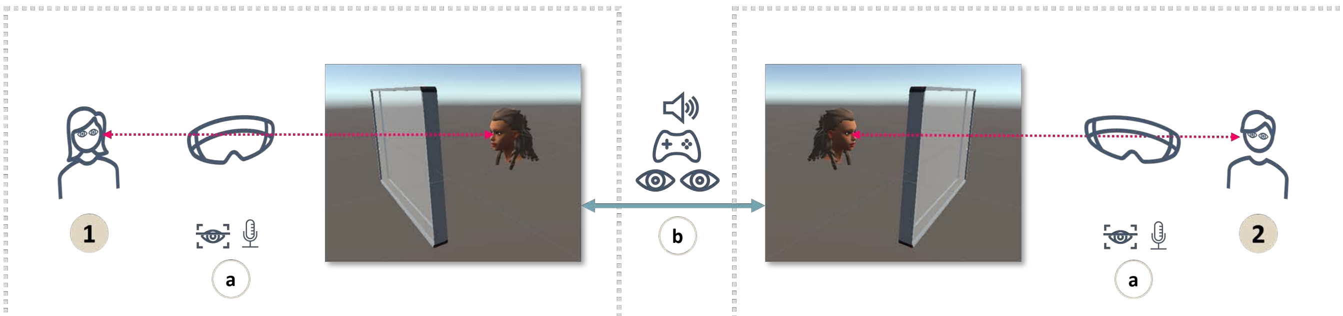

Figure 1: An overview of the GlassBoARd system. The first (1) and second user (2) are wearing Augmented Reality (AR) headsets with built-in eye-trackers and microphones (1a and 2a). Both users see the semi-transparent GlassBoARd with an avatar behind it representing the other user. Gaze, audio, and work-related data (e.g., AR annotations) are transmitted (b). Through the two avatars, GlassboARd allows the users to keep eye contact with each other (indicated through the red lines).

ABSTRACT

Recent research on remote collaboration focuses on improving the sense of co-presence and mutual understanding among the collaborators, whereas there is limited research on using non-verbal cues such as gaze or head direction alongside their main communication channel. Our system – GlassBoARd – permits collaborators to see each other's gaze behavior and even make eye contact while communicating verbally and in writing. GlassBoARd features a transparent shared Augmented Reality interface that is situated in-between two users, allowing face-to-face collaboration. From the perspective of each user, the remote collaborator is represented as an avatar that is located behind the GlassBoARd and whose eye movements are contingent on the remote collaborator's instant eye movements. In three iterations, we improved the design of GlassBoARd and tested it with two use cases. Our preliminary evaluations showed that GlassBoARd facilitates an environment for conducting future user experiments to study the effect of sharing eye gaze on the communication bandwidth.

CHI'24 EA, May 11-16, 2024, Honolulu, HI

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/10.1145/3613905.3650965

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; **Mixed / augmented reality**; **Collaborative interaction**; • **Computing methodologies** → **Perception**.

KEYWORDS

eye tracking, augmented reality, non-verbal cues, remote collaboration, CSCW, gaze, presence

ACM Reference Format:
Kenan Bektaş, Adrian Pandjaitan, Jannis Strecker, and Simon Mayer. 2024. GlassBoARd: A Gaze-Enabled AR Interface for Collaborative Work. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3613905.3650965

1 INTRODUCTION

In recent years, the use of augmented and mixed reality (AR and MR) applications kept gaining attention in the scientific and business discourse. Especially with the diffusion of necessary supporting features (e.g., wireless internet connection, improved sensing capabilities, better form factor, and battery life of the headsets) such applications can today reasonably be used to support shared experiences or computer-supported collaborative work (CSCW).

Studies on the usefulness of AR/MR applications yield mixed results concerning their industrial applicability. Researchers showed that MR can make manufacturing training easier without significant differences in knowledge retention [16], MR-based remote support can also be beneficial in error mitigation [1]. Others, such as [40] argue that sharing awareness cues in MR applications can improve user experience without interfering with the task performance. On the other hand, providing spatial, temporal, and visual realism [24], and overcoming limited interoperability and user acceptance of MR devices remain as challenges to be addressed in various personal or industrial use cases [4, 39].

Besides the aforementioned examples it seems that not many organizations have embedded AR/MR solutions into their day-to-day operations, especially in collaborative work and remote support. However, in our personal and professional lives, concepts such as tele-operation and remote collaboration saw a significant upturn during the COVID-19 pandemic [39, 42]. Thus, in the near future it is highly likely that work and collaboration might shift focus to using AR/MR technologies as one of the main enablers of collaborative operations that must be performed remotely, and for more general CSCW applications. But until this point is reached, many questions about AR/MR and remote collaboration still need to be answered. In this paper, we are concerned with two specific questions that contribute to the scholarship of AR/MR-based collaboration: **Q1:** How can we develop an AR system to improve the remote and collaborative work experiences with non-verbal communication cues? **Q2:** How does sharing eye gaze (as a non-verbal cue) in an AR setting affect the communication bandwidth of remote collaboration?

To address Q1, we collect evidence from a broad review of the scholarship relevant to MR-enabled remote collaboration (see Section 2). This review informs the design of the GlassBoARd system that we introduce in Section 3. GlassBoARd enables collaborators to annotate a shared transparent interface while seeing each others' instant eye movements. In contrast to previous solutions, which employ physical media or virtual reality headsets, GlassBoARd uses an optical see-through AR headset. Conventional video-based telecommunication solutions (e.g., WhatsApp, Zoom, MS-Teams) are limited in allowing users to make actual eye contact due to the disparity between the front camera and the monitor. In these face-to-face settings, existing approaches either imitate the position of users' eyes (e.g., [33] or in NVIDIA Broadcast[1]) or they require computationally demanding solutions [30]. GlassBoARd overcomes these issues by mapping the eye movements of each user to the eye movements of an avatar that represents the respective collaborator (Figure 1). Thus, GlassBoARd allows collaborators to use their actual eye movements as an intentional deictic reference and a non-verbal communication cue. In two collaborative use cases (i.e., assembly of an electronic circuit and playing a board game) we tested the feasibility of transmitting non-verbal cues on the communication bandwidth and the experienced co-presence. In remote collaboration settings, GlassBoARd facilitates a test environment to study the effect of sharing eye gaze on the communication bandwidth. In the future, we will use GlassBoARd as a tool to study Q2 in controlled experiments.

[1]https://www.nvidia.com/en-us/design-visualization/software/broadcast-app/

2 RELATED WORK

Since the 1980s, computer supported cooperative [14] or collaborative work [10] (CSCW) solutions provide multiple users a shared (spatio-temporal) workspace [23] which is useful in remotely performed personal or professional activities [39]. Early CSCW systems allowed sharing data and virtual representations of the users, so that the seams between users and the virtual environment would fade away, allowing them to navigate through data as if it was "malleable space" [12]. This way, users can switch between individual and shared tasks or change their viewpoints to accommodate their need for specific information contexts [12]. Through avatars even complex communication cues such as hand gestures and facial expressions can be conveyed that enrich the feeling of *presence* [12]. Presence is the subjective experience of inhabiting one place even when being physically situated in another [48]. It has been found that an elevated feeling of presence and especially co-presence [11, 19], which applies the concept of presence to social settings (e.g., face-to-face interaction that is mediated in teleconferencing), improves mutual understanding and task performance [28, 37, 38].

2.1 Gaze as a Non-Verbal Communication Cue

While the possibility of verbal (written or spoken) communication is presupposed for most CSCW applications, the transmission of non-verbal communication cues (e.g., hand gestures, head direction, and eye gaze) are at the focus of interest in a few early [45] and some more recent studies [3, 21, 40, 41]. In multiple empirical studies, eye gaze was found as a powerful non-verbal cue that can regulate interpersonal interaction and express intimacy [31] and that can reliably modulate and guide individuals attention towards each other (e.g., in eye contact) or towards jointly attended objects [35]. In comparison to the verbal communication cases, there is limited evidence that reassures the promising potential of such non-verbal cues in collaborative desktop [44] or MR [40] environments. Velichkovsky showed that eye movements can serve as an intentional deictic reference that significantly reduces the bandwidth of verbal communication (i.e., number of spoken words) and increases the efficiency of collaborative problem solving [44]. Therefore, sharing non-verbal cues, such as eye gaze, is an important feature to consider when developing new collaborative MR applications.

2.2 MR-Enabled CSCW Applications

The MR continuum that is proposed by Milgram and Kishino [36] focuses on computer generated visual experiences in real, augmented, and virtual environments. MR-enabled CSCW applications are expected to reduce seams between the real and virtual, allowing users to experience "natural communication behaviors" [9, 17, 20]. In remote meetings or teleconferences, MR allows users to get a feeling "beyond being there" [8], indicating that they can facilitate the sense of co-presence. A seamless MR-enabled CSCW application should allow users to solve tasks faster and communicate more naturally by exhibiting a minimum set of features such as setting virtual objects (and annotations) into a context with real objects and enabling independence (e.g., of viewports) and individuality among the collaborators [9].

In the early 2000s, the cost, complexity, and the sensing and rendering capabilities of the MR hardware posed a central limitation to using them in collaborative settings [2, 7, 9]. Today, the form factor and usability of MR HMDs (e.g., Microsoft Hololens 2, Varjo XR-3, Magic Leap 1 and 2, HTC Vive ProEye, Apple Vision Pro) have significantly improved. They can overlay virtual information on top of real scenes, track the instant 3D position and movements of the user (including their head, hands, and eyes), provide users with novel and intuitive interactions, and they hold the potential to become more than a visual interface [24, 39, 43]. Developers can create convincing MR experiences with the help of sensors embedded to the HMDs including gyroscopes or infrared sensors in connection with base stations[2] to measure position, orientation and acceleration of the user, outward facing cameras to process the environment through a video stream on the device or transmit it to a remote computer and infrared sensors to measure the user's eye gaze.

In real face-to-face settings and in some CSCW applications such as ClearBoard [22], collaborators can see where others are looking by observing their eye movements. However, in most MR settings this poses a substantial challenge because remote collaborators often cannot see each other and in cases where they can, it is often solved just by overlaying a video stream of one user's (e.g., an expert) webcam view as an inset on the field of view of the HMD of their collaborator (e.g., in Microsoft's Dynamics 365 Remote Assist[3]). This does not profit from the contextual meaning to the eye gaze, as it is not in sync with what the remote expert actually sees or looks at. Some research exists on how to mitigate such issues with the help of overlaying a virtual beam or circle onto the field of view of the executing collaborator, indicating where the remote expert is looking at a given time [40]. Findings from recent experiments have shown that MR HMDs can allow collaborators to share their field of view which in return can improve their communication and the sense of togetherness [27], while gaze-enabled HMDs can improve the reading experience in virtual reality (VR) [34] and remote collaboration on physical tasks in MR [25] environments. Collaboration among multiple users can also be facilitated through a transparent display that can additionally allow a direct interaction with objects that are seen through the display [26, 32].

2.3 Implications for the Design of AR/MR CSCW Systems

Towards building AR/MR enabled CSCW systems, what a user sees in their viewport has a large impact on the perception of the system. For example, a study by Kim and colleagues has shown that users prefer to control their own viewports independently and prefer to give an indication of the other viewports [27]. Thus, contrary to what might seem intuitive, what you see is what I see (WYSIWIS) systems do not necessarily create a higher sense of togetherness and awareness [27]. A further advantage of such a setup is that it prevents an unstable video feed [18]. Individual viewports also allow for limiting the display of information only to relevant contexts for a given user [47]. In some MR-enabled CSCW cases, both remote and local users may prefer an expert-and-executing-collaborator setting [46] to a setting where both users were given equal knowledge and responsibilities [27]. Further studies on collaborative work also implied that such a setting might increase general awareness [13] and with it togetherness and eventually the feeling of presence.

As the collaborative work mainly deals with communication and negotiation [12], it is crucial to make communication over AR/MR technology as intuitive and easy as possible. This requires means to exchange most if not all communication cues, such as spoken language, eye gaze cues and hand gestures between the collaborators. Aside from these communication cues, which can naturally be used in non CSCW settings, improving communication is a large potential use case of the technological advantages of AR/MR technology. On the one hand, users seem to expect AR/MR technology to improve their CSCW experience rather than just keeping it as good as it was by "enhancing" reality [8]. On the other hand, since it is able to display virtual (meta) information into a user's field of view, AR/MR technology provides the possibility to add even more context to the communication between the collaborators and meet these expectations. For example, for including visual communication cues, an annotation system can be used to allow a remote expert to add notes into the executing user's field of view [27]. With cameras and suitable sensors on board of the HMD, this could be further refined to make the annotations "world locking" meaning that the remote expert can add annotations directly to real objects in the executing user's surroundings [15, 27]

In conclusion, the design of AR/MR systems to support remote CSCW has to focus extensively on fostering communication. Ways to achieve this are to enable users to independently set their field of view while being shown the viewports of the other person, to have a clear division of work and to also enable the transmission of non-verbal communication cues such as eye gaze. Keeping these points in mind, we developed a system and conducted a user evaluation to examine the capabilities of remote AR/MR collaboration.

2.4 The ClearBoard System and Experiments

Ishii and colleagues designed ClearBoard, which has inspired our work, to study how non-verbal communication cues can be effectively shared in CSCW settings [22]. The ClearBoard is a shared-drawing medium that aims to overcome the spatial, temporal and functional seams (i.e., constraints or limitations) [6] between interpersonal space (IPS) and the shared workspace (SWS) [22]. The SWS is the space where the actual collaborative task takes place and the IPS is described as the space where the communication over the task happens [22]. Losing essential communication cues between SWS and IPS causes difficulties in shifting focus between these spaces and a less efficient communication [22]. Although on a common physical whiteboard it is also not possible to focus on both spaces at the same time, in contrast to most Groupware programs it is possible to switch between these spaces very naturally, for example, by making quick eye contact with the other collaborator [22]. Even today, popular CSCW software does not seem to have solved this problem. For example, Microsoft Teams[4]

[2]https://store.steampowered.com/valveindex
[3]https://dynamics.microsoft.com/en-us/mixed-reality/remote-assist/
[4]https://www.microsoft.com/de-ch/microsoft-teams/group-chat-software

or the conferencing platform Zoom[5] gained rapid popularity during the COVID-19 pandemic. However, they clearly separate SWS (e.g., a video stream of a user's desktop) and IPS (e.g., the video chat) and thereby create an artificial seam. Some of these seams might even be more pronounced now, as Microsoft Teams does not display the other participants of a conference (by default) to the user sharing a PowerPoint presentation, virtually limiting the IPS to audio transmission.

The ClearBoard-0 was a glass screen between two collaborators sitting face-to-face to each other on opposite sides of this screen [22]. The major drawback of this setup was that written letters appeared mirrored to one collaborator. To solve this problem amongst others, two improved versions of the ClearBoard (using a system of networked computers, cameras and projectors) were proposed [22]. In the end, the system consisted of two drawing boards from which the user's own drawings were recorded and on which the video stream of the other collaborator and their drawings were projected [22]. The reported results were that the seams between IPS and SWS actually were reduced and increased eye contact could be detected [22]. This in turn was argued to be potentially crucial for future CSCW applications [22], not least because it might lead to increased awareness and satisfaction with the tool.

The impact of the ClearBoard experiment has been so extensive that when AR/MR/VR technologies became viable, it was already recreated and examined in a virtual setting multiple times. For example, [20] developed a VR tool which allowed co-located or remote collaborators to write on a virtual transparent screen but also to share non-verbal communication cues through their own virtual avatars. The results of this experiment imply that reducing seams in a virtual world is also beneficial to task performance and the feeling of connectedness with other users [20]. Therefore, it can be inferred that the face-to-face setting of the ClearBoard which allowed to share non-verbal communication cues also proved promising for VR applications as it likely improves the feeling of presence. In two experiments, Kiyokawa and colleagues studied the communication behavior of co-located users with optical see-through (AR), stereoscopic, and monoscopic video see-through and VR headsets [29]. Both experiments were conducted in collaborative face-to-face settings. The results of the first experiment showed that real world visibility (as in the AR case) has a significant effect on the communication behavior. The AR case significantly improved collaborative search performance and significantly reduced the need for extra communication (e.g., pointing more than once and using deictic phrases) because of the availability of the non-verbal communication cues (i.e., gaze). Furthermore, participants of this experiment find that compared to the other cases the AR case provides a more natural view and it was easier for them to see where their partner was looking and pointing at. In the second experiment, Kiyokawa and colleagues explored the interplay between the IPS and three different SWSs. The shared workspace (i.e., a two-dimensional virtual grid) was located a) vertically on a *wall* that is off to the side of the users allowing them to see it from similar perspectives, b) horizontally on a *table* between the users, and c) again vertically but in a *floating* position between the users thus allowing them to see it from opposite directions. The results showed that placing the shared workspace between users as in the floating case allows a natural, social, and easy communication. It promotes more active behavior (e.g., use of body language), motivates involvement, and causes less misunderstanding. However, users liked the wall case most because they could see the SWS from similar perspectives.

3 GLASSBOARD SYSTEM

Drawing on the original ClearBoard experiments in combination with the properties of a distributed AR-enabled CSCW system, as introduced above, we developed the GlassBoARd system using the Microsoft HoloLens 2 (HL2) device for a local user who collaborates with a remote expert as shown in Figure 2. To visualize the virtual elements of our experiment setup we created an application with the Unity game engine implementing the Microsoft Mixed Reality Toolkit (MRTK). This allowed us to deploy the app to the HL2 as well as executing it on a PC running Windows 11. Similar to the ClearBoard, we went through multiple iterations in our design.

3.1 First Version

In the first version of GlassBoARd (V1), both users saw a virtual version of ClearBoard on their respective screens (HL2 and PC). In V1, communication from the HL2 to the PC was facilitated by further transmitting a video stream from a stationary camera to be displayed behind the virtual GlassBoARd of the remote user (e.g., an expert as shown on Figure 2). The local user in turn only saw the GlassBoARd with an avatar representing the remote user behind it. To transmit non-verbal gaze cues, we fitted the PC with a Tobii Pro Fusion Eye tracker and accessed the eye tracking functionality of the HL2 to track the eye movements of the users in relation to the GlassBoARd and send them back and forth between each other. The avatars' eyes thus matched the movements of the actual users' eyes. In addition, both the transparent interface of the GlassBoARd as well as the eye movements were mirrored in a way that both users were looking at the same orientation of the interface and of the items that were either placed or drawn on it. If the HL2 users now looked to the left, for instance, the PC user would see the avatar looking to its right, which is the user's left. Communication for both users was therefore possible in written form (on the GlassBoARd, where also images such as the circuit diagrams could be displayed by the remote expert), verbally by talking (even though they were physically separated, the users were within earshot of each other) or through non-verbal means such as the eye gaze from the local to the remote user. We used the same Unity app for the HL2 and the PC, and implemented the mirror-networking package[6] to allow for the interconnection between these apps using only Unity which worked but turned out to be not as easy to maintain and extend.

While testing V1, we employed the repair of a broken circuit as a collaboration scenario, where a remote user (in expert role) gave verbal and visual instructions to a local user. Before the task started, the remote expert on the PC saw the avatar of the local user. After it began, however, the avatar was disabled on the PC and instead the above-mentioned video stream of the workbench was shown. We could observe that this kind of task and setup lead to the majority of the communication being conducted orally only

[5]https://zoom.us/

[6]https://mirror-networking.com

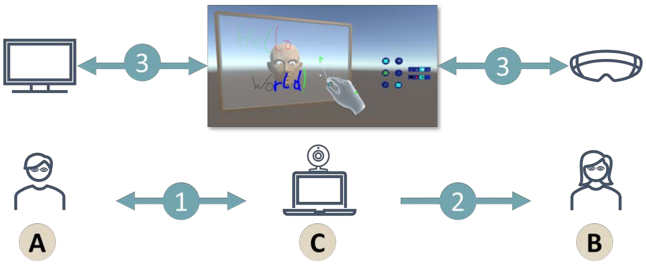

Figure 2: GlassBoARd-V1. The setup allows collaboration between a remote user with a PC (A) and a local user wearing an HL2 (B). The data streams include (1) the WebRTC (audio/video) between the remote user and a laptop computer (C) that records the workspace of the local user (2), and a custom protocol to synchronise game objects (3) of the GlassBoARd-V1 such as the avatar and annotations.

which prompted us to employ a different task which incentivized the use of the non-verbal cues as a core GlassBoARd feature.

This first version thus mostly acted as a proof of concept that our setup could track and transmit eye track gaze cues with acceptable accuracy and that writing on the GlassBoARd was possible, even though some users reported difficulties with the HL2's hand input system of pointing and pinching. Moreover, trying out the system, first users reported that they did not like the appearance of the avatar (see Figure 2). They said that it put them into the metaphorical "uncanny valley", meaning that the anthropomorphic characteristics of the model were pronounced enough to represent a human but just not enough to be convincing that it was a real human but rather a doll. Hence, due to performance problems and the mentioned difficulties regarding future extension, we decided to improve the application to provide it with a better foundry.

3.2 Second Version

In the second version of GlassBoARd (V2), we aimed for a more symmetrical solution with respect to the exchange of communication cues, in which both users see each other represented through avatars. This way, first, the amount of exchanged data is reduced, as no video stream needs to be transmitted, which might be beneficial in settings where bandwidth is limited. Second, it is possible that the feelings of co-presence and togetherness can be further enhanced if both users see an actual representation of their co-worker with corresponding eye movements, instead of just their hands through the video stream. This resulted in us removing the video stream for the remote expert on the Desktop and constantly displaying only the GlassBoARd and the avatar to both users. To counter the problem of the "uncanny valley", in V2, we furthermore replaced the avatar with a friendlier and more human looking model (Figure 3) from Readyplayer Me[7].

Lastly we changed the connection between the HL2 and the Desktop from a direct connection using the Mirror library for Unity to transmitting the gaze and GlassBoARd data formatted as JSON

[7]https://readyplayer.me/avatar

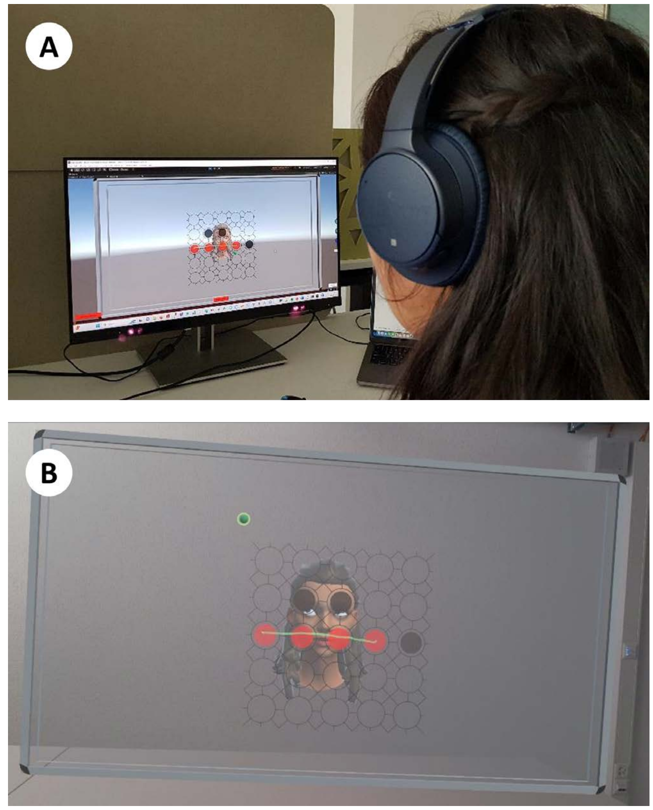

Figure 3: GlassBoARd-V3. The views from the local user's perspective in a desktop setup (A) and a remote user with an HL2 (B), where the green dot represents the instant point of interest of the local user.

via the WebSockets protocol to a central express.js server which relayed it to the Desktop and HL2 applications respectively.

From testing out V2 we concluded that the overall setup was more robust, and the avatars better accepted than it was in the first version. However, we were missing a suitable task to be conducted utilizing the GlassBoARd, as the repair of the broken circuit from V1 would not benefit as much when no video transmission of the workbench was possible any longer.

3.3 Third Version

Inspired by the original ClearBoard experiment where the task consisted of teaching instructions for the game GO, for V3 we implemented a collaboration interface for Teeko (Figure 3), a Japanese board game similar to Four-in-a-row and Tic-Tac-Toe. We modified the express.js-server so that it was able to process the game logic on the board and additionally provided a Web interface through which the conductor of the experiment could easily select the desired experimental condition, start it and log the gaze data.

During the preliminary testing of V3, we placed test users in different rooms when collaborating, so as to create a true distributed setting. In V3, the verbal communication was transmitted over a Zoom call which allowed us to record the communication at both

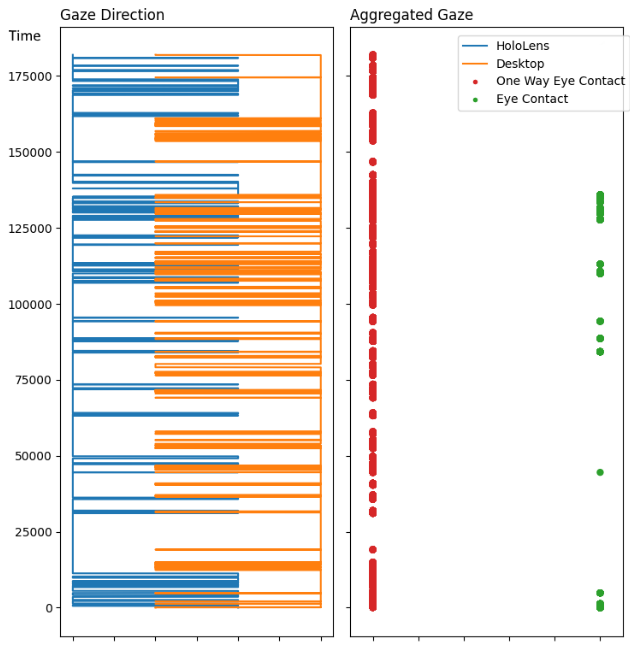

Figure 4: The gaze data evaluation as a function of time (in milliseconds). In the left subplot, overlapping lines mean that both users looked at each others' eyes. In the right subplot, if one user looks at the other on their own, a (red) point is created on the left. The (green) points on the right represent the moments when the users had eye contact.

ends and analyze the effect of eye gaze sharing on the communication bandwidth. We asked test users to bring their personal noise-canceling headphones and laptops. After the tests, we transcribed these recordings and subsequently analyzed the *total number of words* and the *number of words per minute*. As the two participants were connected through a Zoom call, we took advantage of Zoom's capability of recording a separate audio track for each user, as well as one with both participants' speech included. We then cut the recording of each user into individual recordings. While Zoom also offers the possibility to transcribe these recordings, we found that its results required too much manual labour to guarantee correct transcriptions. Therefore we used faster-whisper[8], a re-implemenation of OpenAI's Whisper speech recognition model[9], for transcribing the recordings. Using faster-whisper with Whisper's *large-v2* model produced good results in our setup with little manual correction required. The source code of V3 and the transcription scripts can be found in our public repository.[10]

4 FUTURE WORK

This paper focuses on our introduced Q1: *How can we develop an AR system to improve the remote and collaborative work experiences with non-verbal communication cues?*. The current version of our system (GlassBoARd-V3) has the necessary features to address this question, albeit additional improvements are possible. Since the GlassBoARd was created for the Universal Windows Platform (UWP)[11] in Unity, it can be easily installed on other MR platforms (and HMDs). For instance, the app could be uploaded to two HL2s (instead of one HL2 and one PC) and used to connect two users in a completely MR-enabled CSCW setting. Future research can focus on different configuration of devices.

In the current version, we used a Zoom call in the assessment of the bandwidth of verbal communication. However, in the future, the HL2 or another HMD can be directly used for maintaining the verbal as well as non-verbal communication without depending on video streaming among the collaborators. Furthermore, the eye movement events (e.g., fixations and blinks) can be successfully detected in AR-HMDs [5]. We think that this information can be transferred among the collaborators and rendered on their respective avatars, which in return can improve the perceived sense of co-presence.

Even though MR technologies are starting to gain traction among the general public, our preliminary tests quickly showed that there are still usability issues especially concerning their input methods. In our tests, those users who were not familiar with the HL2's input system of pointing and pinching needed more time to learn placing markers on the GlassBoARd and often ignored their partner avatars' eye movements. Aside from allocating more time for training and familiarizing users with our solution, we should hence make explicit that we are mirroring the eye gaze of the other person.

These observations inform the design of the experiments that we plan to conduct with GlassBoARd. In a next step, we will use GlassBoARd-V3 in a controlled experiment to study our introduced Q2: *How does sharing eye gaze (as a non-verbal cue) in an AR setting affect the communication bandwidth of remote collaboration?*. Specifically, following the promising findings documented in [29, 44] and the recommendations of a recent review of MR-based remote collaboration on physical tasks [46], in the user evaluation of V3, we will focus on realistic collaborative assembly, maintenance, and repair tasks with objects such as lego pieces, jigsaw puzzles, and mechanical parts.

5 CONCLUSIONS

We present GlassBoARd that allows two users to collaboratively annotate a shared transparent AR interface while seeing each others' instant eye movements. GlassBoARd maps the eye movements of each user to the eye movements of an avatar that represents the respective collaborator and allow them to make eye contact. In this way, we provide a solution that facilitate non-verbal communication between the users based on their actual eye movements (instead of imitating the movements of their eyes). Our solution is not computationally costly and might be preferable for preserving the privacy of users in collaborative settings. However, further user evaluation needs to be conducted to support this claim.

ACKNOWLEDGMENTS

This work was funded by the Swiss Innovation Agency Innosuisse (#48342.1 IP-ICT) and the Basic Research Fund of the University of St.Gallen.

[8]https://github.com/guillaumekln/faster-whisper
[9]https://github.com/openai/whisper
[10]https://github.com/Interactions-HSG/GBoARd
[11]https://learn.microsoft.com/de-de/windows/uwp/

REFERENCES

[1] Andrea Aschauer, Irene Reisner-Kollmann, and Josef Wolfartsberger. 2021. Creating an Open-Source Augmented Reality Remote Support Tool for Industry: Challenges and Learnings. *Procedia Computer Science* 180 (2021), 269–279. https://doi.org/10.1016/j.procs.2021.01.164

[2] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. 2001. Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications* 21, 6 (2001), 34–47. https://doi.org/10.1109/38.963459

[3] Huidong Bai, Prasanth Sasikumar, Jing Yang, and Mark Billinghurst. 2020. A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376550

[4] Kenan Bektas, Jannis Rene Strecker, Simon Mayer, and Markus Stolze. 2022. EToS-1: Eye Tracking on Shopfloors for User Engagement with Automation. In *AutomationXP22: Engaging with Automation, CHI'22*. CEUR Workshop Proceedings. http://www.alexandria.unisg.ch/266339/

[5] Kenan Bektaş, Jannis Strecker, Simon Mayer, and Kimberly Garcia. 2024. Gaze-enabled activity recognition for augmented reality feedback. *Computers & Graphics* (2024), 103909. https://doi.org/10.1016/j.cag.2024.103909

[6] Mark Billinghurst. 2003. Return to Reality. In *Proceedings of the 1st International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia (GRAPHITE '03)*. Association for Computing Machinery, New York, NY, USA, 12. https://doi.org/10.1145/604471.604475

[7] Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A Survey of Augmented Reality. *Foundations and Trends® in Human–Computer Interaction* 8, 2-3 (2015), 73–272. https://doi.org/10.1561/1100000049

[8] Mark Billinghurst and Hirokazu Kato. 1999. Collaborative Mixed Reality. In *Mixed Reality*, Yuichi Ohta and Hideyuki Tamura (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 261–284. https://doi.org/10.1007/978-3-642-87512-0_15

[9] Mark Billinghurst and Hirokazu Kato. 2002. Collaborative augmented reality. *Commun. ACM* 45, 7 (July 2002), 64–70. https://doi.org/10.1145/514236.514265

[10] M. Billinghurst, S. Weghorst, and T. Furness. 1998. Shared space: An augmented reality approach for computer supported collaborative work. *Virtual Reality* 3, 1 (March 1998), 25–36. https://doi.org/10.1007/BF01409795

[11] Frank Biocca, Chad Harms, and Judee K. Burgoon. 2003. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators and Virtual Environments* 12, 5 (Oct. 2003), 456–480. https://doi.org/10.1162/105474603322761270

[12] E. F. Churchill and D. Snowdon. 1998. Collaborative virtual environments: An introductory review of issues and systems. *Virtual Reality* 3, 1 (March 1998), 3–15. https://doi.org/10.1007/BF01409793

[13] Paul Dourish and Victoria Bellotti. 1992. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work - CSCW '92*. ACM Press, Toronto, Ontario, Canada, 107–114. https://doi.org/10.1145/143457.143468

[14] Clarence A. Ellis, Simon J. Gibbs, and Gail Rein. 1991. Groupware: some issues and experiences. *Commun. ACM* 34, 1 (Jan. 1991), 39–58. https://doi.org/10.1145/99977.99987

[15] Ali Reza Emami, Gabriel Takacs, Gavin Dean Lazarow, and Skyler Mark Goodell. 2021. Connecting spatial anchors for augmented reality. https://patents.google.com/patent/US20210350612A1/en?oq=Patent+%23:+US20210350612

[16] Mar Gonzalez-Franco, Rodrigo Pizarro, Julio Cermeron, Katie Li, Jacob Thorn, Windo Hutabarat, Ashutosh Tiwari, and Pablo Bermell-Garcia. 2017. Immersive Mixed Reality for Manufacturing Training. *Frontiers in Robotics and AI* 4 (Feb. 2017). https://doi.org/10.3389/frobt.2017.00003

[17] Jens Emil Sloth Grønbæk, Ken Pfeuffer, Eduardo Velloso, Morten Astrup, Melanie Isabel Sønderkær Pedersen, Martin Kjær, Germán Leiva, and Hans Gellersen. 2023. Partially Blended Realities: Aligning Dissimilar Spaces for Distributed Mixed Reality Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–16. https://doi.org/10.1145/3544548.3581515

[18] Pavel Gurevich, Joel Lanir, and Benjamin Cohen. 2015. Design and Implementation of TeleAdvisor: a Projection-Based Augmented Reality System for Remote Collaboration. *Computer Supported Cooperative Work (CSCW)* 24, 6 (Dec. 2015), 527–562. https://doi.org/10.1007/s10606-015-9232-7

[19] Chad Harms and Frank Biocca. 2004. Internal Consistency and Reliability of the Networked Minds Measure of Social Presence. http://cogprints.org/7026/

[20] Zhenyi He, Ruofei Du, and Ken Perlin. 2020. CollaboVR: A Reconfigurable Framework for Creative Collaboration in Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Porto de Galinhas, Brazil, 542–554. https://doi.org/10.1109/ISMAR50242.2020.00082

[21] Zhenyi He and Ken Perlin. 2020. Exploring the Effectiveness of Face-to-face Mixed Reality for Teaching with Chalktalk. *arXiv:1912.03863 [cs]* (Aug. 2020). http://arxiv.org/abs/1912.03863

[22] Hiroshi Ishii, Minoru Kobayashi, and Jonathan Grudin. 1992. Integration of inter-personal space and shared workspace: ClearBoard design and experiments. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work - CSCW '92*. ACM Press, Toronto, Ontario, Canada, 33–42. https://doi.org/10.1145/143457.143459

[23] Hiroshi Ishii and Naomi Miyake. 1991. Toward an open shared workspace: computer and video fusion approach of TeamWorkStation. *Commun. ACM* 34, 12 (Dec. 1991), 37–50. https://doi.org/10.1145/125319.125321

[24] Yuta Itoh, Tobias Langlotz, Jonathan Sutton, and Alexander Plopski. 2022. Towards Indistinguishable Augmented Reality: A Survey on Optical See-through Head-mounted Displays. *Comput. Surveys* 54, 6 (July 2022), 1–36. https://doi.org/10.1145/3453157

[25] Allison Jing, Kunal Gupta, Jeremy McDade, Gun A. Lee, and Mark Billinghurst. 2022. Comparing Gaze-Supported Modalities with Empathic Mixed Reality Interfaces in Remote Collaboration. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Singapore, Singapore, 837–846. https://doi.org/10.1109/ISMAR55827.2022.00102

[26] Shunichi Kasahara, Valentin Heun, Austin S. Lee, and Hiroshi Ishii. 2012. Second surface: multi-user spatial collaboration system based on augmented reality. In *SIGGRAPH Asia 2012 Emerging Technologies*. ACM, Singapore Singapore, 1–4. https://doi.org/10.1145/2407707.2407727

[27] Seungwon Kim, Mark Billinghurst, and Gun Lee. 2018. The Effect of Collaboration Styles and View Independence on Video-Mediated Remote Collaboration. *Computer Supported Cooperative Work (CSCW)* 27, 3-6 (Dec. 2018), 569–607. https://doi.org/10.1007/s10606-018-9324-2

[28] Seungwon Kim, Weidong Huang, Chi-Min Oh, Gun Lee, Mark Billinghurst, and Sang-Joon Lee. 2023. View Types and Visual Communication Cues for Remote Collaboration. *Computers, Materials & Continua* 74, 2 (2023), 4363–4379. https://doi.org/10.32604/cmc.2023.034209

[29] K. Kiyokawa, M. Billinghurst, S.E. Hayes, A. Gupta, Y. Sannohe, and H. Kato. 2002. Communication behaviors of co-located users in collaborative AR interfaces. In *Proceedings. International Symposium on Mixed and Augmented Reality*. IEEE Comput. Soc, Darmstadt, Germany, 139–148. https://doi.org/10.1109/ISMAR.2002.1115083

[30] Jesper Kjeldskov, Jacob H. Smedegård, Thomas S. Nielsen, Mikael B. Skov, and Jeni Paay. 2014. EyeGaze: enabling eye contact over video. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. ACM, Como Italy, 105–112. https://doi.org/10.1145/2598153.2598165

[31] Chris L. Kleinke. 1986. Gaze and eye contact: A research review. *Psychological Bulletin* 100, 1 (1986), 78–100. https://doi.org/10.1037/0033-2909.100.1.78

[32] Katja Krug, Wolfgang Buschel, Konstantin Klamka, and Raimund Dachselt. 2022. CleAR Sight: Exploring the Potential of Interacting with Transparent Tablets in Augmented Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Singapore, Singapore, 196–205. https://doi.org/10.1109/ISMAR55827.2022.00034

[33] Claudia Kuster, Tiberiu Popa, Jean-Charles Bazin, Craig Gotsman, and Markus Gross. 2012. Gaze correction for home video conferencing. *ACM Transactions on Graphics* 31, 6 (Nov. 2012), 1–6. https://doi.org/10.1145/2366145.2366193

[34] Geonsun Lee, Jennifer Healey, and Dinesh Manocha. 2022. VRDoc: Gaze-based Interactions for VR Reading Experience. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Singapore, Singapore, 787–796. https://doi.org/10.1109/ISMAR55827.2022.00097

[35] Kate T. McKay, Sarah A. Grainger, Sarah P. Coundouris, Daniel P. Skorich, Louise H. Phillips, and Julie D. Henry. 2021. Visual attentional orienting by eye gaze: A meta-analytic review of the gaze-cueing effect. *Psychological Bulletin* 147, 12 (Dec. 2021), 1269–1289. https://doi.org/10.1037/bul0000353

[36] Paul Milgram and Fumio Kishino. 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Information Systems* vol. E77-D, no. 12 (Dec. 1994), 1321–1329.

[37] J Mortensen, V Vinayagamoorthy, M Slater, A Steed, B Lok, and M C Whitton. 2002. Collaboration in Tele-Immersive Environments. (2002), 10.

[38] Michael Narayan, Leo Waugh, Xiaoyu Zhang, Pradyut Bafna, and Doug Bowman. 2005. Quantifying the benefits of immersion for collaboration in virtual environments. In *Proceedings of the ACM symposium on Virtual reality software and technology (VRST '05)*. Association for Computing Machinery, New York, NY, USA, 78–81. https://doi.org/10.1145/1101616.1101632

[39] Jason Orlosky, Misha Sra, Kenan Bektaş, Huaishu Peng, Jeeeun Kim, Nataliya Kos'myna, Tobias Höllerer, Anthony Steed, Kiyoshi Kiyokawa, and Kaan Akşit. 2021. Telelife: The Future of Remote Living. *Frontiers in Virtual Reality* 2 (2021). https://www.frontiersin.org/article/10.3389/frvir.2021.763340

[40] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. 2019. The Effects of Sharing Awareness Cues in Collaborative Mixed Reality. *Frontiers in Robotics and AI* 6 (Feb. 2019), 5. https://doi.org/10.3389/frobt.2019.00005

[41] Prasanth Sasikumar, Lei Gao, Huidong Bai, and Mark Billinghurst. 2019. Wearable RemoteFusion: A Mixed Reality Remote Collaboration System with Local Eye Gaze and Remote Hand Gesture Sharing. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, Beijing, China, 393–394. https://doi.org/10.1109/ISMAR-Adjunct.2019.000-3

[42] Thad Starner. 2022. During the COVID-19 Pandemic, Everyone Had a Head-up Display. *XRDS: Crossroads, The ACM Magazine for Students* 29, 2 (Dec. 2022),

42–47. https://doi.org/10.1145/3571301
[43] Jannis Strecker, Khakim Akhunov, Federico Carbone, Kimberly García, Kenan Bektaş, Andres Gomez, Simon Mayer, and Kasim Sinan Yildirim. 2023. MR Object Identification and Interaction: Fusing Object Situation Information from Heterogeneous Sources. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 3, Article 124 (sep 2023), 26 pages. https://doi.org/10.1145/3610879
[44] Boris M. Velichkovsky. 1995. Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics & Cognition* 3, 2 (Jan. 1995), 199–223. https://doi.org/10.1075/pc.3.2.02vel
[45] Roel Vertegaal. 1999. The GAZE groupware system: Mediating joint attention in multiparty communication and collaboration. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '99)*. Association for Computing Machinery, New York, NY, USA, 294–301. https://doi.org/10.1145/302979.303065 Number of pages: 8 Place: Pittsburgh, Pennsylvania, USA.
[46] Peng Wang, Xiaoliang Bai, Mark Billinghurst, Shusheng Zhang, Xiangyu Zhang, Shuxia Wang, Weiping He, Yuxiang Yan, and Hongyu Ji. 2021. AR/MR Remote Collaboration on Physical Tasks: A Review. *Robotics and Computer-Integrated Manufacturing* 72 (2021), 102071. https://doi.org/10.1016/j.rcim.2020.102071
[47] Peng Wang, Shusheng Zhang, Mark Billinghurst, Xiaoliang Bai, Weiping He, Shuxia Wang, Mengmeng Sun, and Xu Zhang. 2020. A comprehensive survey of AR/MR-based co-design in manufacturing. *Engineering with Computers* 36, 4 (Oct. 2020), 1715–1738. https://doi.org/10.1007/s00366-019-00792-3
[48] Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *PRESENCE: Teleoperators & Virtual Environments* 7, 3 (June 1998), 225–240. https://doi.org/10.1162/105474698565686